Linear Regression

Regression analysis is one of the most common methods of prediction. It is used whenever we have a causal relationship between variables. A great of predictive modelling and practice is done through regression analysis. It becomes extremely powerful when complemented by techniques like factor analysis, and one can surely find many academic papers based on it. Moreover, fundamentals of regression analysis are used in supervised machine learning, so we can how regression is a must for data science. The general point is the following. Among other factors, the amount of money you spend depends on the amount of money you earn. In the same way, the amount of time you devote to this course is affected by your motivation to learn additional statistical method. You can qualify these relationships and many others using regression analysis. A typical step by step approach will be used to analyze this topic. Starting with a simple linear regression model along the way, this write-up emphasizes the building of regression, how to interpret it, and how to compare different models. A deep understanding of the fundamentals will be developed to interconnect with the knowledge we have so far. The process goes like this. You get a sample data, design a model that explains the data and then make predictions for the whole population based on the model you developed. There is a dependent variable labelled y and an independent variable labelled x1 x2 and so forth, these are called the predictors. Y is the function of the x variables, and the regression model is the linear approximation of this function. The easiest regression model is the simple linear regression.0  + 1x1 + ε. Where y is the variable we are trying to predict and it is called the dependent variable and x is the independent variable. 0 is a constantand is the intercept of the regression line with y axis. 1 is the slope of the regression line it shows how much y changes for each unit. The distance between the observed value and regression line is the estimate of the error term epsilon ε. It is a point estimate called residual The income a person receives depends on the number of years of education the person has received. The dependent variable is income he principle of **least squares**is one of the popular methods for finding a curve fitting a given data. Say (x1, y1), (x2, y2)….(xn, yn) . As for correlation, it does not imply causation. Correlation measures the degree of relationship between two variables, while regression seeks to highlight how one variable affects the other or what changes it causes to the other correlation does not capture causality but the degree of interrelation between the two variables. Regression is based on causality. It shows no degree of connection but causes an effect. The correlation between x and y is the same as the correlation between y and x. linear regression is known for the best fitting line that goes through the data points and minimizes the distance between them, while correlation is a single point.



A regression model